Tractable Bayesian Estimation of Smooth Transition Vector Autoregressive Models*

Martin Bruns† Michele Piffer‡

September 4, 2023

Abstract

We develop a tractable way of estimating the parameters ruling the nonlinearity in the popular Smooth Transition VAR model, and identify structural shocks using external instruments. This jointly offers an alternative to the option of identifying shocks recursively and calibrating key parameters. In an illustration, we show that monetary policy shocks generate larger effects on economic activity during economic expansions compared to economic recessions. We then document that calibrating rather than estimating the parameters ruling the nonlinearity of the model can lead to values for which the key results are lost. This suggests caution in the calibration of these parameters.

JEL classification: C32, E52.

Keywords: Nonlinear models, Bayesian Econometrics, proxy SVARs, monetary policy shocks.

---

*We are thankful to Anastasia Allayioti, Robin Braun, Annika Camehl, Luca Fanelli, Lutz Kilian, Glen Livingston, Helmut Lütkepohl, Haroon Mumtaz, Edoardo Palombo, Giovanni Pellegrino and Ivan Petrella for helpful comments and suggestions. Michele Piffer is thankful for the financial support received from the European Union's Horizon 2020 research and innovation program, Marie Sklodowska-Curie grant agreement number 744010. Martin Bruns thanks the German Academic Scholarship Foundation for financial support.

†University of East Anglia, UK, School of Economics.

‡Corresponding author: King's Business School, King's College London, Bush House, 30 Aldwych, London WC2B 4BG, UK. `e-mail:` m.b.piffer@gmail.com

1 Introduction

Smooth Transition Vector Autoregressive models are frequently used in applied Macroeconomics to study if the effects of macroeconomic shocks change depending on the state of the economy. Popular applications in the literature include Weise (1999), who studies monetary policy shocks, and Auerbach and Gorodnichenko (2012), who study fiscal policy shocks.[1]

In this paper we show that the practice of calibrating (rather than estimating) key parameters in Smooth Transition models should be reconsidered, and we offer a tractable way forward. In Smooth Transition (ST hereafter) models, the transition across states is pinned down by two parameters: c, which captures the midpoint of the underlying transition variable along which the nonlinearity is explored (say, a measure of how close the economy is to a recession), and γ, which captures the speed at which the model adjusts across regimes. It is common practice in part of the literature to set $c = 0$ and to calibrate γ to match a subjective definition of recessions.[2] However, moving across equally plausible calibrations of (γ, c) can have dramatic effects on the results. No obvious alternative solution is currently present in the literature. Methods of estimating (γ, c) have been proposed both in the frequentist and in the Bayesian framework, see for example Granger and Teräsvirta (1993) and Gefang and Strachan (2010), respectively. However, the literature is still exploring how to best combine such methods with the identification of structural shocks, which in this model is usually achieved using the recursive identification.

The model used in our framework builds on the existing Bayesian literature. It

[1]Smooth transition models were independently introduced by Bacon and Watts (1971), Goldfeld and Quandt (1972) and Maddala (1977) in univariate regression models. They were then extended to univariate autoregressive models by Chan and Tong (1986) and Teräsvirta (1994) and made popular in a multivariate framework by Auerbach and Gorodnichenko (2012). For discussions of the smooth transition model, see Granger and Teräsvirta (1993), Teräsvirta et al. (2010) and van Dijk et al. (2002).

[2]This approach is followed by Auerbach and Gorodnichenko (2012) and several subsequent papers, including Bachmann and Sims (2012), Berger and Vavra (2014), Caggiano et al. (2015), and Caggiano et al. (2022).

implements two straightforward modifications that suffice to generate a convenient improvement in the tractability of the model. (1) We identify structural shocks using external instruments, and do so following Paul (2020) by adding the instruments as regressors in a STVAR model. Since this modelling device allows for a rich set of structural shocks to generate time-varying responses on impact, we argue that a fully heteroskedastic structure in the model is less needed. The same is not true in the traditional STVAR framework without instruments as regressors, where homoskedasticity implies that no structural shock can generate time-varying effects already on impact. We hence use a homoskedastic model as our baseline specification, and propose a heteroskedastic extension to assess the robustness of the results. (2) We then leverage on the fact that a conditionally conjugate prior allows for the derivation of the marginal (rather than the conditional) posterior distribution of (γ, c). This simplifies posterior sampling for two reasons. First, because visual inspection of the marginal posterior can aid the researcher by uncovering potential posterior sampling challenges even before running the algorithm. Second, because the marginal distribution can be sampled using a standard Metropolis-Hastings algorithm.

The computational advantage of working with the marginal rather than the conditional distribution has been exploited in univariate applications of ST autoregressive models, see Bacon and Watts (1971) and Lubrano (2000). To the best of our knowledge, we are the first to extend their univariate approach to a multivariate structural model. In STVARs, existing contributions sample from the *conditional* posterior distribution of (γ, c), which by construction changes as the sampler proceeds over the conditioning parameters. Accordingly, preliminary visual inspection is less viable, and sampling requires the computationally more demanding Block-Metropolis-Hastings-within-Gibbs sampler (as in Gefang and Strachan, 2010, Gefang, 2012, Galvão and Owyang, 2018, Laumer and Philipps, 2020). We validate our procedure using Monte Carlo simulations, which show that a homoskedastic STVAR model with an instrument as regressor can still successfully estimate the impulse responses of a heteroskedastic

STVAR model.

We are aware of only one contribution that uses instruments to identify structural shocks in a STVAR model. Rather than controlling for instruments are regressors, Carriero et al. (2018) specify the STVAR as in Auerbach and Gorodnichenko (2012) and identify a monetary shock via instruments using the linear methods by Mertens and Ravn (2013) and Stock and Watson (2012). While their approach is more general than ours in terms of the heteroskedasticity of the model, it implies that the impact impulse responses can change across regimes only up to a constant, a restriction which is not present in our approach. This restriction is, in our view, worth relaxing, since it does not fit naturally within the rich type of heteroskedasticity that Carriero et al. (2018) allow for in the reduced form specification of their model. See also Bolboaca and Fischer (2019) for an earlier (frequentist) use of STVAR models combined with restrictions on the forecast error variance decomposition.

We use the model to revisit the long lasting debate on whether monetary policy shocks generate larger or smaller effects in economic downturns or during expansions. The literature provides contradicting results on this topic: Weise (1999) and Santoro et al. (2014) found that monetary policy shocks have a larger effect on the real economy during recessions, while Tenreyro and Thwaites (2016) and Alpanda et al. (2021) found the opposite result. Our framework, which estimates (γ, c), suggests that monetary policy shocks have a larger effect on real GDP in economic expansions. We use this illustration to study what happens if one calibrates rather than estimates (γ, c). We show that the results are largely lost when using different equally plausible calibrations of (γ, c). We then show that calibrating $c = 0$ effectively imposes symmetry in how much time the model is allowed to spend in what the model labels as recessionary compared to expansionary periods. By contrast, in allowing the data to deliver a negative estimate for c, our model reveals that only few periods in the sample are interpreted as deep, yet short recessions (see also Hamilton, 1989 for a similar result).

The paper is organized as follows. Section 2 outlines the model and discusses inference in a general framework. Section 3 shows the application to monetary policy shocks and discusses how results change when calibrating rather than estimating (γ, c). Section 4 concludes.

2 The model

The tractability of the multivariate ST framework used in our paper is due to the particular way in which we specify the model and select prior beliefs.

2.1 The smooth transition structural VAR model

In its general form, the model is given by

$$\begin{aligned} \boldsymbol{y}_t = & \quad g_{t-1} \cdot \big(\Pi_a \boldsymbol{x}_{t-1} + B_a \boldsymbol{m}_t + D_a \boldsymbol{q}_t\big) + \\ & + (1 - g_{t-1}) \cdot \big(\Pi_b \boldsymbol{x}_{t-1} + B_b \boldsymbol{m}_t + D_b \boldsymbol{q}_t\big) + \boldsymbol{u}_t, \end{aligned} \tag{1a}$$

$$\boldsymbol{u}_t \sim N(\boldsymbol{0}, \Sigma_t), \tag{1b}$$

$$g_{t-1} = g(z_{t-1}, \gamma, c) = \frac{1}{1 + e^{-\gamma(z_{t-1} - c)}}, \quad \gamma > 0, \tag{1c}$$

with Σ_t given by one of the following three options:

$$\Sigma_t = \Sigma, \tag{2a}$$

$$\Sigma_t = g_{t-1} \cdot \Sigma_a + (1 - g_{t-1}) \cdot \Sigma_b, \tag{2b}$$

$$\Sigma_t = h_{t-1} \cdot \Sigma, \quad \text{with} \quad h_{t-1} = g_{t-1} + e^{\psi} \cdot (1 - g_{t-1}), \quad \psi \geq 0. \tag{2c}$$

The $k \times 1$ vector $\boldsymbol{y}_t$ contains the endogenous variables of the model. The $kp \times 1$ vector $\boldsymbol{x}_{t-1} = (\boldsymbol{y}_{t-1}, ..., \boldsymbol{y}_{t-p})'$ contains the p lags of the variables of the model. The $k_m \times 1$ vector $\boldsymbol{m}_t$ includes observable external instruments for the structural shocks of interest. The $k_q \times 1$ vector $\boldsymbol{q}_t$ potentially contains control variables, a constant, and external measures of other structural shocks. $\boldsymbol{m}_t$ and $\boldsymbol{q}_t$ are exogenous in the model. Σ_t can be modelled as homoskedastic (equation 2a), heteroskedastic according to a smooth transition (2b), or heteroskedastic up to a time-varying scalar h_{t-1} (2c). $g(z_{t-1}, \gamma, c)$ indicates a scalar transition logistic function, which evolves endogenously if z_t is part of $\boldsymbol{y}_t$ or is a function of it, and is exogenous otherwise. Model (1)-(2) is effectively a time-varying linear combination of two latent processes,

$$\boldsymbol{y}_t = \Pi_i \boldsymbol{x}_{t-1} + B_i \boldsymbol{m}_t + D_i \boldsymbol{q}_t + \boldsymbol{u}_t^i, \qquad i = a, b, \tag{3}$$

which are associated with the extreme values $g(z_{t-1}, \gamma, c) = 1$ and $g(z_{t-1}, \gamma, c) = 0$. As the transition variable z_{t-1} evolves over time, the logistic transition function $g(z_{t-1}, \gamma, c)$ varies and implies a different point within the continuum of linear combinations of models (3) (van Dijk et al., 2002, Teräsvirta et al., 2014).

The general specification from model (1) helps appreciate how the approach followed in our paper differs from the existing literature. Assume for the moment that no instruments are introduced in the regressions, hence structural shocks only drive $\boldsymbol{y}_t$ via the reduced form innovations $\boldsymbol{u}_t$. Replacing Π with $\Pi_t = g_{t-1} \cdot \Pi_a + (1 - g_{t-1}) \cdot \Pi_b$ in the linear VAR model $\boldsymbol{y}_t = \Pi \boldsymbol{x}_{t-1} + \boldsymbol{u}_t$ allows for nonlinear impulse responses for horizons $h \geq 1$, without raising technical challenges in the estimation. A homoskedastic model (equation 2a) keeps the estimation manageable, but assumes that impulse responses are linear for $h = 0$. To allow for nonlinear impulse responses on impact, several contributions introduce heteroskedasticity of the type from equation (2b). This is the standard approach to STVAR models. It is used either for reduced form analysis or combined with the recursive identification of the shocks, and it is used by either calibrating or estimating the parameters (γ, c).[3]

The approach followed in our paper differs from the traditional approach outlined above in two straightforward but crucial ways, which jointly make the approach

[3]See Auerbach and Gorodnichenko (2012) and Caggiano et al. (2022) for models that calibrate (γ, c), and Gefang and Strachan (2010) and Galvão and Owyang (2018) for methods that estimate (γ, c).

particularly tractable for applied research. (1) We follow Paul (2020), who identifies structural shocks using instruments, but uses them as regressors following a VARX approach, rather than the two-step Proxy SVAR approach by Stock and Watson (2012) and Mertens and Ravn (2013). Paul (2020) shows an equivalence between linear VARXs and linear Proxy-SVARs. He then shows that a time-varying VARX à la Primiceri (2005) is very tractable, while a time-varying Proxy SVAR remains technically challenging. We follow Paul (2020) and introduce instruments $\boldsymbol{m}_t$ for the shocks of interest, which allows the impulse responses to the shocks captured by $\boldsymbol{m}_t$ to potentially change on impact over time. (2) By introducing not only instruments for the shocks of interest but also for other shocks (via $\boldsymbol{q}_t$), we allow for a rich set of structural shocks to potentially generate time-varying impact effects. This partly makes heteroskedasticity from equation (2b) less compelling, because even the homoskedastic version of the model allows the effects of several shocks to be nonlinear already on impact, via the instruments. This is not the case under the traditional approach to STVARs, where structural shocks only enter the model via the reduced form innovations.

For the above reasons, we either assume homoskedasticity or heteroskedasticity up to a constant h_{t-1}, equations (2a) and (2c), respectively. It is natural to expect that inference is much more tractable under (2a) or (2c) than under assumption (2b), as we outline more precisely in the next section. We argue that this approach to STVAR models provides a tractable yet flexible approach to study nonlinearities. The detailed Monte Carlo simulation exercises in the Online Appendix, which build in part on Bruns and Lütkepohl (2022), provide encouraging results in favour of our strategy. They show that a homoskedastic model that adds as regressor only an instrument for the shock of interest and imposes $\Sigma_t = \Sigma$ can correctly estimate the true impulse responses even when the data are generated from a model that features $\Sigma_t = g_{t-1} \cdot \Sigma_a + (1 - g_{t-1}) \cdot \Sigma_b$. In addition, as we document in Section 3.3, calibrating rather than estimating (γ, c) can have dramatic effects on the results. This generates demand for a tractable estimation procedure for (γ, c), which we achieve thanks to the combination of instruments as regressors and constraints to heteroskedasticity.

The number of instruments used in the model and the assumptions on Σ_t have implications for the assumptions on the identification of the shocks. As an example, the baseline specification of the application in Section 3 identifies a monetary shock controlling for the instruments of four additional shocks via $\boldsymbol{q}_t$ in a homoskedastic framework. It hence assumes that only up to five structural shocks can potentially generate time-varying effects on impact. The hereroskedastic extension (via 2c) allows the remaining structural shocks, in $\boldsymbol{u}_t$, to generate effects that change on impact up to a constant, which potentially further improves the model's fit to the data. In general, the shocks of interest can be identified either by assuming that each entry of $\boldsymbol{m}_t$ correlates only with one shock of interest, or by allowing for one or more instruments to correlate with potentially more than one shock. In the latter case, the VARX approach can, in principle, disentangle this set of shocks using sign restrictions on the difference in the impulse responses. It cannot, however, introduce restrictions on the correlations of the instruments with the shocks, as done in the linear Proxy SVAR framework, because this correlation is not specified (Piffer and Podstawski, 2018, Arias et al., 2021, Giacomini et al., 2022). We leave it for future research to explore how the model from our paper can be combined with sign restrictions.

2.2 Inference

As widely discussed, for instance, in Lubrano (2001), van Dijk et al. (2002), Gerlach and Chen (2008), and Livingston and Nur (2017), inference in ST models is made challenging by the way in which (γ, c) affect the transition variable and the likelihood function of the model. For $\gamma = 0$, $g(z_{t-1}, \gamma, c) = 0.5$ for any value of z_{t-1}, making (Π_a, B_a, D_a) not separately identified from (Π_b, B_b, D_b). Values of γ in the neighbourhood of zero make the parameters $(\Pi_a, B_a, D_a, \Pi_b, B_b, D_b)$ weakly identified, raising problems in the numerical routines of both frequentist and Bayesian methods, due to the model approaching perfect multicollinearity. As γ increases, $g(z_{t-1}, \gamma, c)$ approaches the indicator function $I[z_{t-1} > c]$, and becomes very steep around the inflection point $z_{t-1} = c$ already for relatively low values of γ. This, in turn, makes any difference in high values of γ hard to detect.

We follow Lubrano (2001), Lopes and Salazar (2006) and Livingston and Nur (2017) and address the above challenge using a Bayesian approach. The parameters of the model are $(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi)$, with $\boldsymbol{\theta} = \text{vec}([\Pi_a, B_a, D_a, \Pi_b, B_b, D_b])$ and $\text{vec}(\cdot)$ the operator that stacks the columns of a matrix vertically. Conditioning on (γ, c, ψ), the model is linear. This makes it natural to restrict the prior on $(\boldsymbol{\theta}, \Sigma)$ to the conditionally conjugate Normal-inverse-Wishart prior. By contrast, the prior on (γ, c, ψ) can be freely selected by the researcher. In short, we use prior beliefs

$$p(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi) \propto \mathrm{I}\{\boldsymbol{\theta}\} \cdot \tilde{p}(\boldsymbol{\theta}|\Sigma, \gamma, c, \psi) \cdot \tilde{p}(\Sigma|\gamma, c, \psi) \cdot \tilde{p}(\gamma, c, \psi), \tag{4a}$$

$$\tilde{p}(\boldsymbol{\theta}|\Sigma, \gamma, c, \psi) = N\Big(\boldsymbol{\mu}, V \otimes \Sigma\Big), \tag{4b}$$

$$\tilde{p}(\Sigma|\gamma, c, \psi) = iW(S, d), \tag{4c}$$

$$\tilde{p}(\gamma, c, \psi) = \text{free.} \tag{4d}$$

The indicator function $\mathrm{I}\{\boldsymbol{\theta}\}$ equals one when $\boldsymbol{\theta}$ satisfies the restrictions of interest (if any), for instance restrictions on the stationarity of the model. While not made explicit in the notation, the hyperparameters $(\boldsymbol{\mu}, V, S, d)$ can be a function of (γ, c, ψ).

The combination of the model specification from Section 2.1 and prior beliefs from equation (4) makes inference particularly tractable, because the joint posterior distribution satisfies

$$p(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi|Y) \propto \mathrm{I}\{\boldsymbol{\theta}\} \cdot \tilde{p}(\boldsymbol{\theta}|Y, \Sigma, \gamma, c, \psi) \cdot \tilde{p}(\Sigma|Y, \gamma, c, \psi) \cdot \tilde{p}(\gamma, c, \psi|Y), \tag{5a}$$

$$\tilde{p}(\boldsymbol{\theta}|Y, \Sigma, \gamma, c, \psi) = N\Big(\boldsymbol{\mu}^*, \bar{V}^*(\gamma, c, \psi) \otimes \Sigma\Big), \tag{5b}$$

$$\tilde{p}(\Sigma|Y, \gamma, c, \psi) = iW(S^*(\gamma, c, \psi), d^*), \tag{5c}$$

$$\tilde{p}(\gamma, c, \psi|Y) \propto \tilde{p}(\gamma, c, \psi) \cdot |\det(S^*(\gamma, c, \psi))|^{-\frac{d+T}{2}} \cdot \left|\det(\bar{V}^{-1} + W(\gamma, c)W(\gamma, c)')\right|^{-\frac{k}{2}}, \tag{5d}$$

with $(\boldsymbol{\mu}^*, \bar{V}^*, S^*, d^*)$ reported in Appendix A. The crucial feature of the joint posterior is the analytical derivation of the marginal posterior $\tilde{p}(\gamma, c, \psi|Y)$ rather than the conditional posterior $\tilde{p}(\gamma, c, \psi|Y, \boldsymbol{\theta}, \Sigma)$. The joint posterior distribution $p(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi|Y)$ can then be sampled in three main steps: (1) use a Metropolis-Hastings algorithm to sample from $\tilde{p}(\gamma, c, \psi|Y)$ (or from $\tilde{p}(\gamma, c|Y)$ under homoskedasticity); (2) conditioning on (γ, c, ψ), draw $(\boldsymbol{\theta}, \Sigma)$ from $\tilde{p}(\Sigma|Y, \gamma, c, \psi)$ and $\tilde{p}(\boldsymbol{\theta}|Y, \Sigma, \gamma, c, \psi)$ using direct sampling; (3) store $(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi)$ if $\boldsymbol{\theta}$ satisfies the restrictions (if any) modelled via the indicator function $\mathrm{I}\{\boldsymbol{\theta}\}$. When calibrating rather than estimating (γ, c, ψ), only Steps 2-3 are needed. We refer to the Online Appendix for a detailed discussion of the algorithm.[4]

The key feature that allows for both the tractable estimation of (γ, c) and the identification via external instruments is the use of a Normal prior for $\boldsymbol{\theta}$ with a Kronecker structure on the variance (i.e. $\boldsymbol{\theta} \sim N\big(\boldsymbol{\mu}, V \otimes \Sigma\big)$), rather than introducing prior independence of $\boldsymbol{\theta}$ from Σ (i.e. $\boldsymbol{\theta} \sim N\big(\boldsymbol{\mu}, \tilde{V}\big)$). In a linear VAR, allowing for prior independence is often desirable, because it is less restrictive and it still allows for efficient sampling via a Gibbs sampler (see for instance Koop and Korobilis, 2010). However, prior independence of $\boldsymbol{\theta}$ on Σ becomes computationally costly in a ST model, where posterior sampling also involves (γ, c), as well as ψ if the model is heteroskedastic. The Bayesian STVAR literature that estimates (γ, c) tends to use an independent Normal prior on $\boldsymbol{\theta}$, which implies sampling the joint posterior distribution using a

[4]The derivations of the posterior distribution are available on the authors' webpage.

Metropolis-Hastings-within-Gibbs sampler (Gefang and Strachan, 2010, Gefang, 2012 and Carriero et al., 2018). The Metropolis-Hastings step is required for $\tilde{p}(\gamma, c|Y, \boldsymbol{\theta}, \Sigma)$, a distribution which can be ill-shaped in a way that constantly changes as the chain progresses on $(\boldsymbol{\theta}, \Sigma)$. By contrast, imposing a Kronecker structure in the prior for $\boldsymbol{\theta}$, while more restrictive, allows for the analytical derivation of the *marginal* distribution $\tilde{p}(\gamma, c, \psi|Y)$ (or $\tilde{p}(\gamma, c|Y)$ under homoskedasticity), which is constant in $(\boldsymbol{\theta}, \Sigma)$. Hence, no Gibbs sampler is required. In addition, in a homoskedastic setting one can use visual inspection of $\tilde{p}(\gamma, c|Y)$ prior to running the algorithm to better tackle the potentially challenging shape of $\tilde{p}(\gamma, c|Y)$. This allows detecting possible identification problems for (γ, c) prior to running long posterior chains, saving time in the estimation of the model. We illustrate this point graphically in the application in Section 3, see Figure 1.[5]

We conclude this section with a final remark on the inference for γ. For any values of (γ, c) the model is linear in $\boldsymbol{w}_t(\gamma, c) = \left[g_{t-1}\boldsymbol{x}_{t-1}, g_{t-1}\boldsymbol{m}_t, g_{t-1}\boldsymbol{q}_t, (1-g_{t-1})\boldsymbol{x}_{t-1}, (1-g_{t-1})\boldsymbol{q}_t, (1-g_{t-1})\boldsymbol{q}_t\right]$, which are gathered in the matrix $W = [\boldsymbol{w}_1(\gamma, c), .., \ \boldsymbol{w}_T(\gamma, c)]$. The model is not identified for values of γ close to zero, since they make $W(\gamma, c)$ approach singularity, pushing $|\det(\bar{V}^{-1} + W(\gamma, c)W(\gamma, c)')|$ in equation (5d) towards infinity. Contrary to Gefang and Strachan (2010), we found that this inconvenient feature of the model cannot always be offset via the curvature and truncation introduced with $\tilde{p}(\gamma)$. We hence follow Lubrano (2000)'s proposal in the context of univariate ST models. We introduce prior dependence in $(\boldsymbol{\theta}, \gamma)$ and calibrate $V(\gamma)$ such that $\tilde{p}(\boldsymbol{\theta}|\Sigma, \gamma)$ becomes more informative as γ approaches zero. We do this

[5] The computational convenience of building the posterior sampler on the marginal rather than the conditional posterior distribution of selected parameters is already fully acknowledged in threshold models, both in univariate and in multivariate specifications (see Geweke and Terui, 1993 and Forbes et al., 1999, respectively). When Bacon and Watts (1971) extended the threshold model to a univariate ST model, they used the same insight, employing a preliminary visual inspection of the marginal posterior distribution of (γ, c) to guide posterior sampling. To the best of our knowledge, the literature on ST models has followed this approach only within univariate models (see also Lubrano, 2000 and Bauwens et al., 2000, Chapter 8).

by setting $\bar{V}(\gamma) = \begin{pmatrix} \bar{\bar{V}}(\gamma) & 0 \\ 0 & \bar{\bar{V}}(\gamma) \end{pmatrix}$ with $\bar{\bar{V}}(\gamma) = \mathrm{diag}\big(\bar{\bar{\boldsymbol{v}}}'_{\Pi}(\gamma), \bar{\bar{\boldsymbol{v}}}'_{mq}(\gamma)\big)$ with $\underset{kp\times 1}{\bar{\bar{\boldsymbol{v}}}'_{\Pi}(\gamma)} = \Big(\lambda(\gamma)\cdot\boldsymbol{\iota}_k \quad \lambda(\gamma)/2\cdot\boldsymbol{\iota}_k \quad \cdots \quad \lambda(\gamma)/p\cdot\boldsymbol{\iota}_k\Big)$, $\bar{\bar{\boldsymbol{v}}}'_{mq}(\gamma) = \lambda_1 \cdot 100,000 \cdot \boldsymbol{\iota}_{k_m+k_q}$ and $\boldsymbol{\iota}_g$ the unitary vector of dimensions $g \times 1$. Lastly, we model prior dependence in $(\boldsymbol{\theta}, \gamma)$ in a smooth transition fashion as

$$\lambda(\gamma) = \frac{\lambda_1}{1 + e^{-\gamma_\gamma\cdot(\gamma - c_\gamma)}}. \tag{6}$$

In the baseline specification, we set $\lambda_1 = 2$, $\gamma_\gamma = 2$ and $c_\gamma = 1$ so that the prior on $\boldsymbol{\theta}$ is uninformative for values of γ above 1, and progressively but rapidly becomes more informative as γ approaches 1 and decreases further. This ensures that $\big|\bar{V}^{-1}(\gamma) + W(\gamma,c)W(\gamma,c)'\big|$ in $\tilde{p}(\gamma, c|Y)$ does not diverge to infinity as γ approaches zero. See Lubrano (2000) for a similar prior specification in a univariate framework.

3 Monetary policy shocks and the business cycle

We apply the model from Section 2 to study how the effects of a conventional monetary policy shock change depending on the phase of the business cycle at the time when the shock hits the economy. We use this application to illustrate the possible implications of calibrating rather than estimating the parameters (γ, c).

3.1 Data, baseline model specification, priors

The model introduces five monthly US variables: the one-year government interest rate, the measure of spread computed by Gilchrist and Zakrajšek (2012) (GZ Spread), the log of the S&P500 index, the log of interpolated real GDP, and the log of the interpolated GDP deflator. The selection of the variables follows Jarociński and Karadi (2020), except that we replace the GZ Excess Bond Premium with the GZ Spread.[6]

[6]We introduce the GZ Spread rather than the Excess Bond Premium because the former is measured directly on the data. The latter, instead, is the outcome of an estimation from a linear model,

The sample period starts in 1979M7, as in Gertler and Karadi (2015). We extend it until 2020M3, the month when the World Health Organization declared the COVID-19 pandemic. We include 12 lags in the model, as in Gertler and Karadi (2015) and Jarociński and Karadi (2020). The baseline specification assumes homoskedasticity by setting $\psi = 0$.

Our nonlinear application requires specifying the transition variable z_{t-1}. We use a seven-month backward-looking moving average of the quarter-to-quarter log difference of real GDP. We construct the transition variable on quarter-to-quarter variations. We then select seven as the length of the moving average process using the Deviance Information Criterion (DIC) by Spiegelhalter et al. (2002). By construction, higher values of the transition variable are associated with higher growth of real GDP.

Identification of the monetary shock is achieved using the high-frequency instrument based on fed funds futures computed by Gertler and Karadi (2015), which starts in 1990M2 and was updated by Jarociński and Karadi (2020) to cover the period until 2016M11. Jarociński and Karadi (2020) argue that the instrument correlates with a central bank information channel shock, which is believed to generate opposite effects on the financial variables, economic activity and inflation compared to the monetary policy shock of interest. For this reason, we use the "poor man's" version of their instrument, which coincides with their original instrument in the periods when the surprise on the fed funds futures and on the surprise in the S&P500 take the opposite sign, and zero in the remaining periods. We control for the TED spread shock by Stock and Watson (2012), the oil demand and oil supply shocks estimated by Baumeister and Hamilton (2019), and the instrument for uncertainty shocks by Piffer and Podstawski (2018).

Appendix C provides a detailed illustration of the prior distributions used. We

which is at odds with our use of a nonlinear model to explore the dynamics of the economy. The sensitivity analysis shows that the results remain unchanged when using the Excess Bond Premium. The results partly become weaker in recession when using industrial production and CPI inflation rather than real GDP and the GDP deflator, as in Gertler and Karadi (2015). We refer to Appendix B for information on the data source and for how interpolations were implemented.

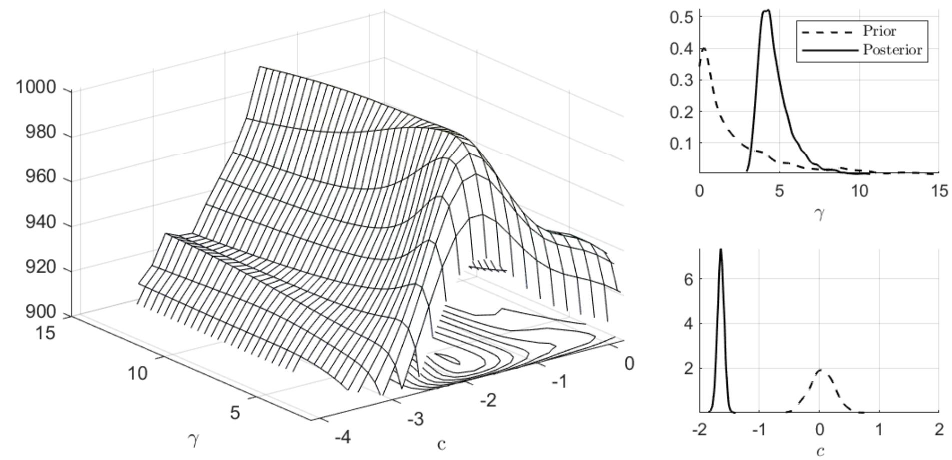
follow common practice and standardize z_{t-1} to make (γ, c) scale-free. We start from a gamma prior for $\tilde{p}(\gamma)$, calibrating its hyperparameters to ensure an expected value equal to 2 and standard deviation equal to 3. We then truncate $\tilde{p}(\gamma)$ to $\gamma \geq 0.01$ to ensure that we stay away from the non-identified region, as also suggested by Gefang and Strachan (2010) and Livingston and Nur (2017). As for $\tilde{p}(c)$, we start from a Normal distribution centred at the median value of z_{t-1}, truncate it to have positive mass only between the 1st and 99th percentiles of z_{t-1}, and calibrate the variance to ensure that 95% probability mass of the truncated prior is between the 30th and 70th percentiles of z_{t-1}. We then set $\tilde{p}(\Sigma|\gamma, c) = \tilde{p}(\Sigma)$, calibrating the hyperparameters as in Kadiyala and Karlsson (1997). Lastly, we set $\tilde{p}(\boldsymbol{\theta}|\Sigma, \gamma, c) = \tilde{p}(\boldsymbol{\theta}|\Sigma, \gamma)$, calibrating the hyperparameters such that the prior is relatively uninformative conditioning on values of γ above 1, and progressively more informative as γ approaches zero. We use the indicator function $\mathrm{I}\{\boldsymbol{\theta}\}$ from equation (4) to restrict both conditionally linear models (3) to be stationary. Since $\tilde{p}(\Sigma)$ requires a training sample, we employ the period 1979M7-1990M1 as a training sample and the remaining sample 1990M2-2020M3 as the estimation sample.[7] The sampler only takes a few minutes to run on a computer with an Intel i7-7700K 4.2GHz Quad Core processor and 64 GB RAM.

3.2 Inference and Results

A key advantage of our model setup is the possibility of graphically exploring the marginal posterior over (γ, c) prior to running the sampler. Figure 1 (left panel) reports this distribution, evaluated numerically over a grid along (γ, c). The figure shows that this distribution is well behaved, suggesting that the Metropolis-Hastings sampler is likely to perform well. Having run the sampler, the right panel compares the

[7]The beginning of the estimation sample coincides with the first period in which the instrument by Jarociński and Karadi (2020) is available. We hence omit it in the model estimated in the training sample, while still including $\boldsymbol{q}_t$, which are available since the beginning of the training sample. In the estimation sample, we set $(m_t, \boldsymbol{q}_t)$ equal to zero for the periods in which the instruments are not available. In Proxy SVAR models, identification can be achieved using a subset of the sample period employed for the estimation of the model. This is not possible in the VARX.

Figure 1: Marginal posterior distribution $\tilde{p}(\gamma, c|Y)$ (in log) and updating of γ, c

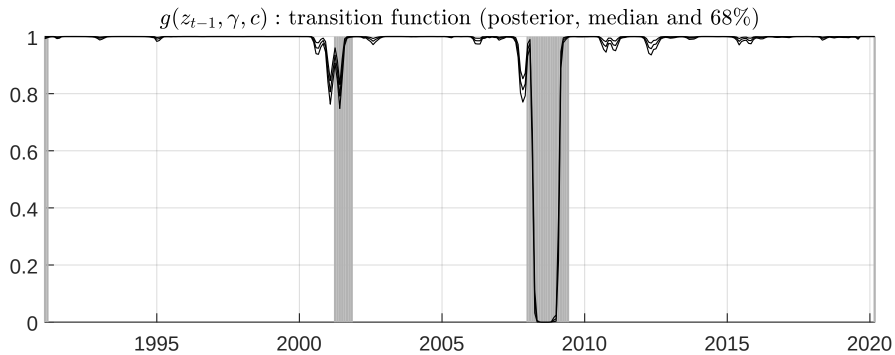


Note: Marginal posterior $\tilde{p}(\gamma, c|Y)$ (left panel). Prior and posterior of γ, c (right panel).

marginal prior distributions to the marginal posterior distributions computed over the posterior draws. The figure shows that indeed the priors are updated and the posterior is centered around the global maximum of the marginal posterior. The update over (γ, c) pins down the posterior of the transition function $g(z_{t-1}, \gamma, c)$, which pins down the transition pattern of the model (see Figure 2). In estimating c at a large and negative value, our model detects that the data are better explained by several periods of mild to strong expansions, and relatively few periods of strong and abrupt recessions. This result is also in line with Hamilton (1989), who shows that the economy is best characterized by sharp and less prolonged periods of recession rather than a symmetric unfolding of similar sized recessions and expansions.

We use the model to study if the impulse responses to a monetary policy shock differ depending on whether the central bank intervenes in a period closer to a recession or closer to an expansion. We build our analysis on nonlinear impulse responses in the spirit of Koop et al. (1996). We use the expressions "recession" and "expansion" to refer to different subperiods of the sample. Our definition of the nonlinear impulse responses defines a recession as all periods in which the transition variable z_{t-1} is below the lowest quartile of all sample values of z_{t-1}, and defines an expansion using

Figure 2: Smooth transition pattern



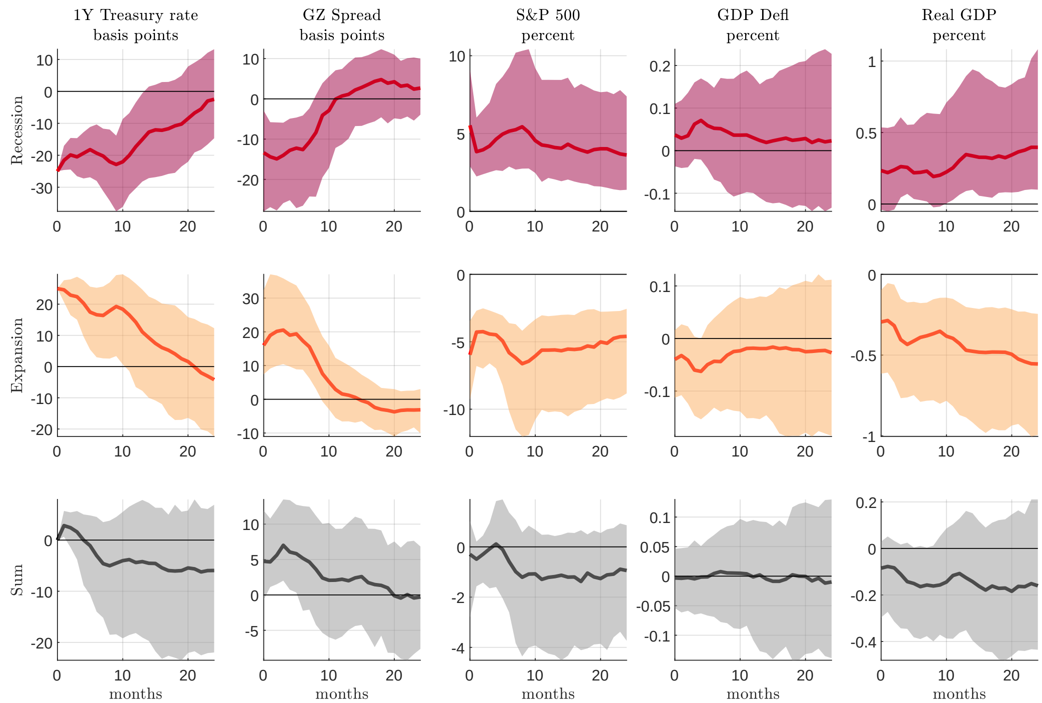
Note: The shaded areas indicate NBER recessions.

the highest quartile. Figure 2 shows that what our model labels as recessions align well with the NBER definitions. We generate a monetary policy shock by scaling the appropriate entry of m_t in equation (1a) to decrease the interest rate on impact by 25 basis points in a recession, or to increase it by the same amount in an expansion. We do so to avoid discussing a less natural monetary expansion in an expansion, or a monetary contraction in a recession.[8] The nonlinear impulse responses average out the future shocks that hit the economy after the initial impact, as well as the actual period at which the shock is simulated, within the recessionary or expansionary subperiod.

As shown in Figure 3, an exogenous monetary expansion that decreases the interest rate by 25 basis points during a recession has a quantitatively strong impact effect on the stock market, as the S&P500 increases by 5.2%. Inflationary pressures are weak, but the effects on output materialize already in the first months after the shock. By contrast, a monetary contraction during an expansionary period generates impulse responses that are, in general, of the opposite sign as the ones in recession, but differ from them qualitatively and quantitatively. While the effects on the GDP deflator is fairly symmetric, the shock has a stronger effect on real GDP in an expansion. A 25 basis point increase in the interest rate in an economic expansion decreases real

[8]As we document in the Online Appendix, the nonlinearity documented in the analysis is driven by the different periods in which the shocks are generated, rather than by the fact that the generated shocks are of opposite sign.

Figure 3: Posterior nonlinear impulse responses



Note: The figure shows pointwise median and 68% bands, using the first definition of the nonlinear impulse responses.

GDP by up to 0.5%, contrary to a 0.3% increase in a recession following a 25 basis point decrease in the interest rate. This is consistent with the fact that both financial variables in the model respond more in expansion than in recession.

Table 1: Strength in the nonlinearities of the nonlinear impulse responses

	horizons				
	0 : 2	3 : 5	6 : 8	9 : 11	12 : 18
One-year interest rate	0.775	0.385	0.260	0.260	0.200
GZ Spread	0.775	0.830	0.650	0.525	0.390
S&P500	0.120	0.325	0.170	0.190	0.080
GDP Deflator	0.380	0.360	0.480	0.480	0.395
Real GDP	0.190	0.125	0.145	0.190	0.155

Notes: For each variable of the model we select a set of the response horizons (e.g. from 0 to 2, from 3 to 5 etc.) and report the share of total posterior draws for which the sum of the impulse responses in recession and expansion is positive in each of the selected horizons.

To assess the strength of the nonlinearities, for each posterior draw we compute the pointwise sum of the impulse responses between recessions and expansions, see Table 1. We first select a set of horizons of the impulse responses, for instance the first quarter from the shock, including the impact effect (this is the column titled 0:2). We then report the percentage of posterior draws for which the sum of the recessionary and expansionary impulse responses are above zero for all horizons within this set. Under linearity, the impulse responses are symmetric, implying that the sum should be centred around zero and the reported probability should equal 50%. As documented in Table 1, we find that the posterior distribution attaches around 80% probability that the GZ Spread is more responsive in an expansion than in an recession within 5 months from the shock. Similarly, we find that the S&P500 responds more in an expansion, particularly on impact, where the probability that the effect is stronger in an expansion is as high as 88%. While weaker, we find evidence that the interest rate reverts back to zero faster in an expansion. Lastly, there is approximately an 85% posterior probability that real GDP responds more in a expansion. This difference is detected at all horizons studied in the analysis. A graphical illustration of Table 1 is shown in last row of Figure 3.

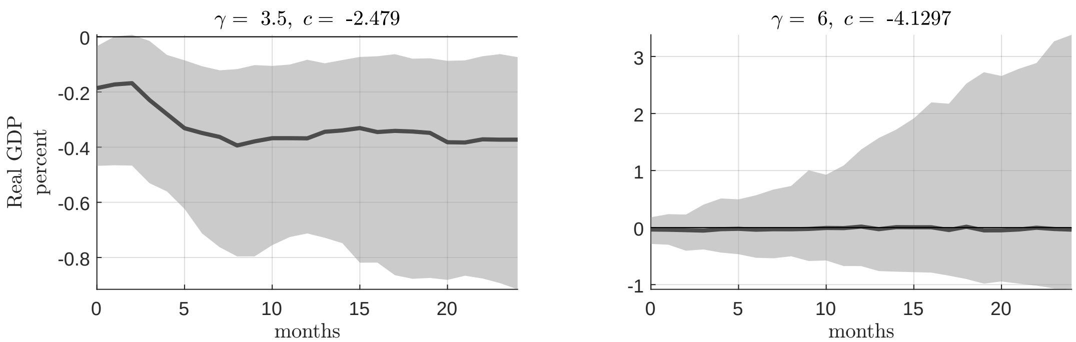
We refer the reader to the Online Appendix, where we document that the results are robust to a wide range of alternative specifications of both the model and the prior beliefs, including the use of a heteroskedastic model and the use the instruments by Romer and Romer (2004) and by Miranda-Agrippino and Ricco (2021).

3.3 Calibrating rather than estimating (γ, c)

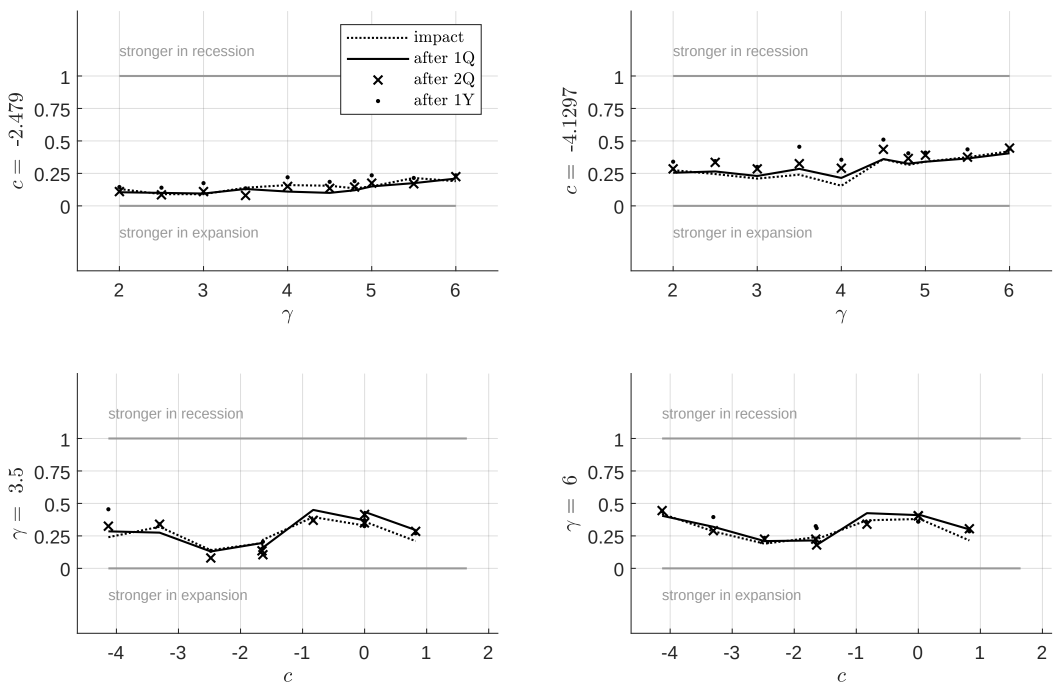
Our estimation delivers the marginal posterior means $E(\gamma|Y) = 4.7959$ and $E(c|Y) = -1.6390$. To study the sensitivity of the results when calibrating rather than estimating (γ, c), we calibrate (γ, c) over a grid with values of γ between 2 and 6 and for values of c between the 1st and 99th percentile of the standardized transition variable (–4.1297 and 1.6479, respectively). The top row of Figure 4 (which should be compared to the last row of Figure 3) reports the sum of the nonlinear responses of real GDP for

Figure 4: Grid search when calibrating rather than estimating (γ, c) (1/2)

(a) Sum of nonlinear IRFs



(b) Probability that monetary policy shocks have a larger effect in recession



Note: See the Online Appendix for the full set of impulse responses associated with the top row of this figure.

two illustrative points of the grid. The baseline results hold even more strongly for the calibration $(\gamma, c) = (3.5, -2.479)$. However, the alternative calibration of $(\gamma, c) = (6, -4.1297)$ shows that the result is largely lost, in that the posterior probability over the sum of the impulse responses is symmetric around 0. This suggests that the results can change considerably for different values of (γ, c). The remaining rows in Figure 4 report the posterior probability that the effects on real GDP are stronger in a recession or in an expansion. It does so by comparing different horizons on the impulse responses for different values of (γ, c). The plots differ for whether they fix c and change γ or vice versa. When holding c fixed, changing the calibration for γ does not substantially change the results (the lines in the middle row are approximately horizontal). However, changing c does have strong effects on the results. This suggests that the differences documented in the top row of Figure 4 are driven by the different values of c rather than of γ.

Figure 5 helps demonstrate why different values of c can have strong effects on the results. Define a period t as recession if $g(z_t, \gamma, c) < g^R$ and as expansion if $g(z_t, \gamma, c) > g^E$, given the subjective threshold values (g^R, g^E) (in the illustration we use $g^R = 0.20$ and $g^E = 0.8$). Conditioning on a value for c, an increase in γ increases the number of periods that the model spends in recessions. This observation is often employed as a justification to calibrate γ to match a target number of periods that the economy spends in recession.[9] However, increasing γ also increases the number or periods that the model associates with expansions. The central plot in the top row of Figure 5 reports the proportion of the sample periods that are labelled as recessions or as expansions for $c = 0$ as γ increases. The fact that the distribution of z_t is slightly skewed to the left implies that $c = 0$ always leads to relatively more periods labelled as expansions relative to recessions for *any* value of γ. Hence, adjusting only γ and holding c fixed effectively introduces a symmetry between how long the model spends in recessionary or expansionary periods. By contrast, modifying both γ *and* c

[9]See, for instance, Auerbach and Gorodnichenko (2012) and several papers that followed their approach.

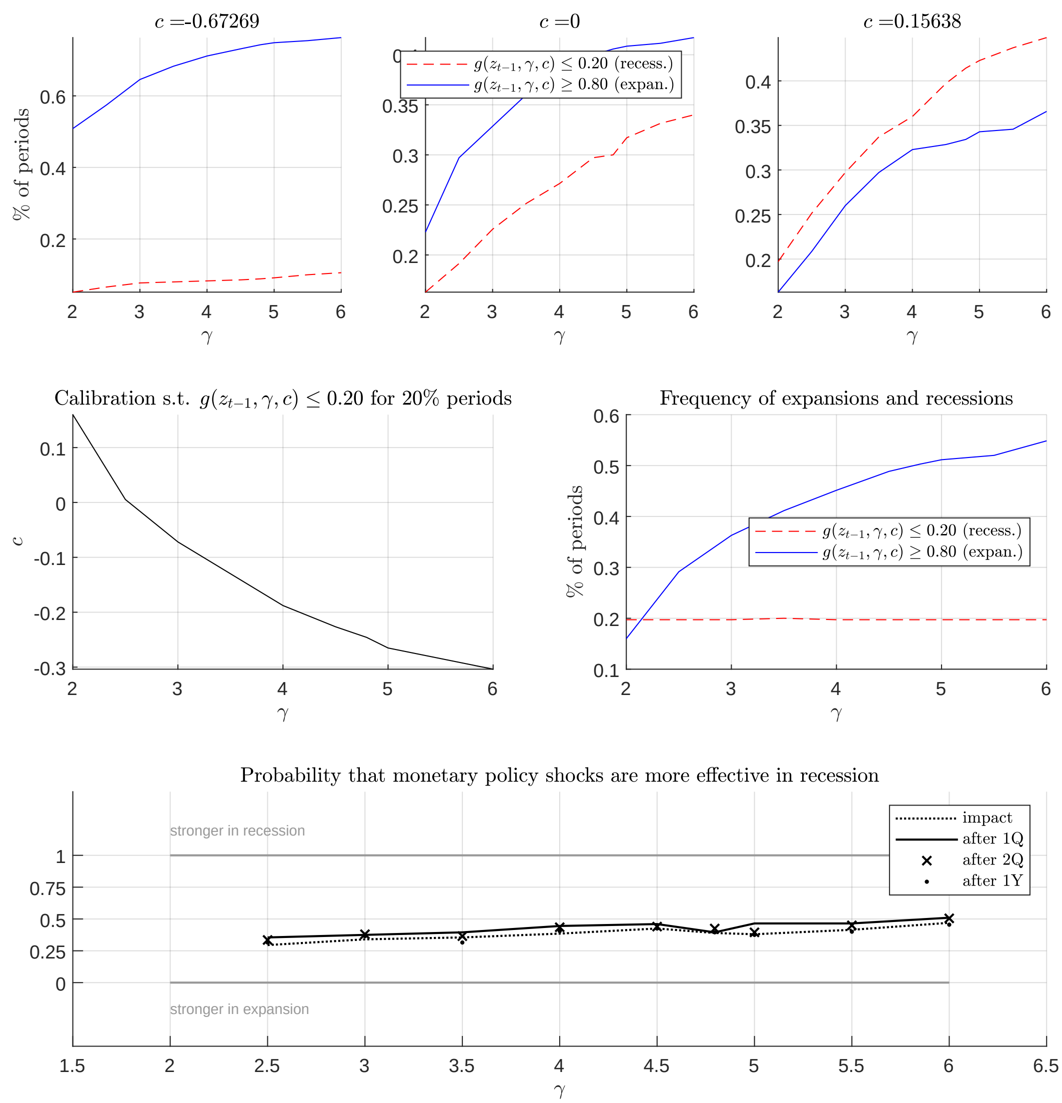
allows the model to spend more time in recessions or in expansions in very different proportions, as documented by the remaining plots in the top row of Figure 5.

The central row of Figure 5 further documents this property of the model by considering different values of γ and calibrating c each time to ensure that the model labels 20% of the sample period as recessionary. For higher γ, lower values of c are required for the number of recessionary periods to remain unchanged. While the different points in the calibration of (γ, c) are equally plausible from the standpoint of common practice in the calibration of γ, they imply very different dynamics for the model, because increasing values of γ increase the proportion of periods labelled as expansion (see also Canepa and Chini, 2016). The last row shows that if we choose different values of γ, while at the same time computing c such that the model spends the desired number of periods in recessions, then we would conclude that monetary policy shocks do not generate asymmetric effects on real GDP. By contrast, by estimating (γ, c), our approach lets the data determine the relative proportion of periods for which the model interprets a selected period as recessionary or as expansionary, without relying on subjective values for (g^R, g^E).

4 Conclusions

This paper develops a tractable way of identifying Smooth Transition VAR models via external instruments while estimating rather than calibrating the key parameters in the smooth transition. The Bayesian approach we pursue in this paper makes the use of instruments for the identification of structural shocks tractable, thereby offering a valid alternative to the recursive identification with calibrated key parameters. We then revisit the linear model on monetary policy shocks by Gertler and Karadi (2015) and Jarociński and Karadi (2020) by letting it interact with the quarter-on-quarter growth in real GDP. Results suggest that monetary policy shocks generate a larger effect on real GDP in expansions rather than in recessions, and that plausible but

Figure 5: Grid search when calibrating rather than estimating (γ, c) (2/2)



Note: See the Online Appendix for the transition function associated with the central and bottom rows of the figure. The first point on the bottom graph is missing due to the model not satisfying the stationarity restrictions for that point of the grid.

arbitrary calibrations of (γ, c) can fail to uncover this result from the data.

References

Alpanda, S., E. Granziera, and S. Zubairy (2021). State dependence of monetary policy across business, credit and interest rate cycles. *European Economic Review 140*, 103936.

Arias, J. E., J. F. Rubio-Ramirez, and D. F. Waggoner (2021). Inference in Bayesian Proxy-SVARs. *Journal of Econometrics 225*(1), 88–106.

Auerbach, A. J. and Y. Gorodnichenko (2012). Measuring the output responses to fiscal policy. *American Economic Journal: Economic Policy 4*(2), 1–27.

Bachmann, R. and E. R. Sims (2012). Confidence and the transmission of government spending shocks. *Journal of Monetary Economics 59*(3), 235–249.

Bacon, D. W. and D. G. Watts (1971). Estimating the transition between two intersecting straight lines. *Biometrika 58*(3), 525–534.

Baumeister, C. and J. D. Hamilton (2019). Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review 109*(5), 1873–1910.

Bauwens, L., M. Lubrano, and J.-F. Richard (2000). *Bayesian inference in dynamic econometric models*. Oxford University Press.

Berger, D. and J. Vavra (2014). Measuring how fiscal shocks affect durable spending in recessions and expansions. *The American Economic Review 104*(5), 112–115.

Bolboaca, M. and S. Fischer (2019). News shocks: Different effects in boom and recession?

Bruns, M. and H. Lütkepohl (2022). Comparison of local projection estimators for proxy vector autoregressions. *Journal of Economic Dynamics and Control 134*, 104277.

Caggiano, G., E. Castelnuovo, V. Colombo, and G. Nodari (2015). Estimating fiscal multipliers: News from a non-linear world. *The Economic Journal 125*(584), 746–776.

Caggiano, G., E. Castelnuovo, and G. Nodari (2022). Uncertainty and monetary policy in good and bad times: A replication of the vector autoregressive investigation by bloom (2009). *Journal of Applied Econometrics 37*(1), 210–217.

Canepa, A. and E. Z. Chini (2016). Dynamic asymmetries in house price cycles: A generalized smooth transition model. *Journal of Empirical Finance 37*, 91–103.

Carriero, A., A. B. Galvao, and M. Marcellino (2018). Credit conditions and the asymmetric effects of monetary policy shocks. *EMF Research Papers 17*.

Chan, K. S. and H. Tong (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis 7*(3), 179–190.

Chow, G. C. and A. Lin (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics 53*(4), 372–375.

Forbes, C. S., G. R. Kalb, and P. Kofhian (1999). Bayesian arbitrage threshold analysis. *Journal of Business & Economic Statistics 17*(3), 364–372.

Galvão, A. B. and M. T. Owyang (2018). Financial stress regimes and the macroeconomy. *Journal of Money, Credit and Banking 50*(7), 1479–1505.

Gefang, D. (2012). Money-output causality revisited–a Bayesian logistic smooth transition VECM perspective. *Oxford Bulletin of Economics and Statistics 74*(1), 131–151.

Gefang, D. and R. Strachan (2010). Nonlinear impacts of international business cycles on the UK: A Bayesian smooth transition VAR approach. *Studies in Nonlinear Dynamics and Econometrics 14*(1), 1–33.

Gerlach, R. and C. W. Chen (2008). Bayesian inference and model comparison for asymmetric smooth transition heteroskedastic models. *Statistics and Computing 18*(4), 391–408.

Gertler, M. and P. Karadi (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics 7*(1), 44–76.

Geweke, J. and N. Terui (1993). Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis 14*(5), 441–454.

Giacomini, R., T. Kitagawa, and M. Read (2022). Robust Bayesian inference in proxy SVARs. *Journal of Econometrics 228*(1), 107–126.

Gilchrist, S. and E. Zakrajšek (2012). Credit spreads and business cycle fluctuations. *American Economic Review 102*(4), 1692–1720.

Goldfeld, S. and R. Quandt (1972). *Nonlinear Methods in Econometrics.* North-Holland, Amsterdam.

Granger, C. W. and T. Teräsvirta (1993). *Modelling non-linear economic relationships.* Oxford University Press.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 357–384.

Jarociński, M. and P. Karadi (2020). Deconstructing monetary policy surprises: the role of information shocks. *American Economic Journal: Macroeconomics 12*(2), 1–43.

Kadiyala, K. R. and S. Karlsson (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics 12*(2), 99–132.

Koop, G. and D. Korobilis (2010). Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics 3*(4), 267–358.

Koop, G., M. H. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics 74*(1), 119–147.

Laumer, S. and C. Philipps (2020). Government spending between active and passive monetary policy.

Livingston, G. and D. Nur (2017). Bayesian inference for smooth transition autoregressive (STAR) model: A prior sensitivity analysis. *Communications in Statistics-Simulation and Computation 46*(7), 5440–5461.

Lopes, H. F. and E. Salazar (2006). Bayesian model uncertainty in smooth transition autoregressions. *Journal of Time Series Analysis 27*(1), 99–117.

Lubrano, M. (2000). Bayesian analysis of nonlinear time-series models with a threshold. In W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjostheim and A. W. Wurtz. (Ed.), *Nonlinear Econometric Modeling in Time Series: Proceedings of the Eleventh International Symposium in Economic Theory*. Cambridge University Press.

Lubrano, M. (2001). Smooth transition GARCH models: A Bayesian perspective. *Recherches Economiques de Louvain/Louvain Economic Review 67*(3), 257–287.

Maddala, G. S. (1977). *Econometrics*. Mc-Graw-Hill.

McCracken, M. W. and S. Ng (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*(4), 574–589.

Mertens, K. and M. O. Ravn (2013). The dynamic effects of personal and corporate income tax changes in the United States. *American Economic Review 103*(4), 1212–1247.

Miranda-Agrippino, S. and G. Ricco (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics 13*(3), 74–107.

Paul, P. (2020). The time-varying effect of monetary policy on asset prices. *Review of Economics and Statistics 102*(4), 690–704.

Piffer, M. and M. Podstawski (2018). Identifying uncertainty shocks using the price of gold. *The Economic Journal 128*(616), 3266–3284.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies 72*(3), 821–852.

Romer, C. D. and D. H. Romer (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review 94*(4), 1055–1084.

Santoro, E., I. Petrella, D. Pfajfar, and E. Gaffeo (2014). Loss aversion and the asymmetric transmission of monetary policy. *Journal of Monetary Economics 68*, 19–36.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series b (Statistical Methodology) 64*(4), 583–639.

Stock, J. H. and M. W. Watson (2012). Disentangling the channels of the 2007-2009 recession. *Brookings Papers on Economic Activity*, 81–135.

Tenreyro, S. and G. Thwaites (2016). Pushing on a string: US monetary policy is less powerful in recessions. *American Economic Journal: Macroeconomics 8*(4), 43–74.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association 89*(425), 208–218.

Teräsvirta, T., D. Tjostheim, C. W. Granger, et al. (2010). *Modelling nonlinear economic time series*. Oxford University Press.

Teräsvirta, T., Y. Yang, et al. (2014). Specification, estimation and evaluation of vector smooth transition autoregressive models with applications. Technical report, Department of Economics and Business Economics, Aarhus University.

van Dijk, D., T. Teräsvirta, and P. H. Franses (2002). Smooth transition autoregressive models—a survey of recent developments. *Econometric Reviews 21*(1), 1–47.

Weise, C. L. (1999). The asymmetric effects of monetary policy: A nonlinear vector autoregression approach. *Journal of Money, Credit and Banking 31*(1), 85–108.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies 21*(4), 1455–1508.

Appendix A: Posterior

The joint posterior of the model is given by

$$
\begin{aligned}
p(\boldsymbol{\theta}, \Sigma, \gamma, c, \psi|Y) &\propto \mathrm{I}\{\boldsymbol{\theta}\} \cdot \tilde{p}(\boldsymbol{\theta}|Y, \Sigma, \gamma, c, \psi) \cdot \tilde{p}(\Sigma|Y, \gamma, c, \psi) \cdot \tilde{p}(\gamma, c, \psi|Y), \\
\tilde{p}(\boldsymbol{\theta}|Y, \Sigma, \gamma, c, \psi) &= N\Big(\boldsymbol{\mu}^*(\gamma, c, \psi), \bar{V}^*(\gamma, c, \psi) \otimes \Sigma\Big), \\
\tilde{p}(\Sigma|Y, \gamma, c, \psi) &= iW\big(S^*(\gamma, c, \psi), d^*\big),
\end{aligned}
$$

$$
\tilde{p}(\gamma, c, \psi|Y) \propto \tilde{p}(\gamma, c, \psi) \cdot |\det(S^*(\gamma, c, \psi))|^{-\frac{d+T}{2}} \cdot \Big|\det(\bar{V}^*(\gamma, c, \psi))\Big|^{\frac{k}{2}} \cdot \Big(\prod_{t=1}^{T} h(z_{t-1}, \gamma, c, \psi)\Big)^{-\frac{k}{2}},
$$

$$
\begin{aligned}
\text{with} \quad \boldsymbol{\mu}^*(\gamma, c, \psi) &= \big(\bar{V}^*(\gamma, c, \psi)\bar{V}(\gamma)^{-1} \otimes I_k\big)\boldsymbol{\mu} + \Big(\bar{V}^*(\gamma, c, \psi)W(\gamma, c)H(\gamma, c, \psi)^{-1} \otimes I_k\Big)\tilde{\boldsymbol{y}}, \\
\bar{V}^*(\gamma, c, \psi) &= \Big[\bar{V}(\gamma)^{-1} + W(\gamma, c)H(\gamma, c, \psi)^{-1}W(\gamma, c)'\Big]^{-1}, \\
S(\gamma, c, \psi)^* &= S + YH(\gamma, c, \psi)^{-1}Y' + N\bar{V}(\gamma)^{-1}N' +
\end{aligned}
$$

$$
-\big(\bar{V}(\gamma)^{-1}N' + W(\gamma, c)H(\gamma, c, \psi)^{-1}Y'\big)'\bar{V}^*(\gamma, c, \psi)\big(\bar{V}(\gamma)^{-1}N' + W(\gamma, c)H(\gamma, c, \psi)^{-1}Y'\big),
$$

$$
d^* = d + T.
$$

with $Y = [\boldsymbol{y}_1, .., \boldsymbol{y}_t, .., \boldsymbol{y}_T]$, $\tilde{\boldsymbol{y}} = \text{vec}(Y)$ and $W(\gamma, c)$ discussed at the end of Section 2.2. In the special case of homoskedasticity ($\psi = 0$) the posterior distribution simplifies to

$$\begin{aligned}
\tilde{p}(\boldsymbol{\theta}|Y, \Sigma, \gamma, c) &= N\Big(\boldsymbol{\mu}^*(\gamma, c), \bar{V}^*(\gamma, c) \otimes \Sigma\Big), \\
\tilde{p}(\Sigma|Y, \gamma, c) &= iW\big(S^*(\gamma, c), d^*\big), \\
\tilde{p}(\gamma, c|Y) &\propto \tilde{p}(\gamma, c) \cdot |\det(S^*(\gamma, c))|^{-\frac{d+T}{2}} \cdot \left|\det\big(\bar{V}(\gamma)^{-1} + W(\gamma, c)W(\gamma, c)'\big)\right|^{-\frac{k}{2}},
\end{aligned}$$

$$\begin{aligned}
\text{with} \quad & \boldsymbol{\mu}^*(\gamma, c) = \big(\bar{V}^*(\gamma, c)\bar{V}(\gamma)^{-1} \otimes I_k\big)\boldsymbol{\mu} + \Big(\bar{V}^*(\gamma, c)W(\gamma, c) \otimes I_k\Big)\tilde{\boldsymbol{y}}, \\
& \bar{V}^*(\gamma, c) = \Big[\bar{V}(\gamma)^{-1} + W(\gamma, c)W(\gamma, c)'\Big]^{-1}, \\
& S(\gamma, c)^* = S + YY' + N\bar{V}(\gamma)^{-1}N' + \\
& \qquad - \big(\bar{V}(\gamma)^{-1}N' + W(\gamma, c)Y'\big)'\bar{V}^*(\gamma, c)\big(\bar{V}(\gamma)^{-1}N' + W(\gamma, c)Y'\big), \\
& \quad d^* = d + T.
\end{aligned}$$

Appendix B: Data

Table 2 provides detailed information on data source and transformations. Most of the variables are downloaded from the dataset by McCracken and Ng (2016), 2021/09 vintage. The measure of the spread is downloaded from the Federal Reserve Board of Governors website.[10] The S&P500 is measured as the end of the month closing value, as in Welch and Goyal (2008), updating their series using the CRSP-Centre for Research in Security Press dataset in WRDS (mnemonic i0003). Interpolations are done using the methodology by Chow and Lin (1971) on the monthly series of industrial production (for real GDP) and on the monthly series of both the consumer and the producer price index (for the GDP deflator). The data for the interpolation are downloaded from Federal Reserve Bank of St Louis (mnemonics GDPC1, GDPDEF,

[10] https://www.federalreserve.gov/econresdata/notes/feds-notes/2016/recession-risk-and-the-excess-bond-premium-20160408.html

INDPRO, CPIAUCSL, WPSFD49207). All variables are demeaned before entering the model.

Table 2: Data

	Source	Mnemonics	Transform
GDP Defl	interpolation		log
GZ Spread	FED Board of Governors		-
One-year interest rate	McCracken and NG (2016)	GS1	-
Real GDP	interpolation		log
S&P 500	CRSP dataset in WRDS	i0003	log

Notes: real variables are computed using PCEPILFE, as in Stock and Watson (2012).

Appendix C: Prior Parameters used in the monetary policy application

The hyperparameters for the prior on Σ, $\Sigma \sim iW(S, d)$, are set as in Kadiyala and Karlsson (1997). We set $d = k + 3$, then use the training sample to estimate σ_i as the standard deviation of the residual in a univariate AR(p) process on variable i, and set $S = (d-k-1) \cdot \mathrm{diag}\big(\sigma_1^2, .., \sigma_i^2, .., \sigma_k^2\big)$. For the prior on $\boldsymbol{\theta}$, $\boldsymbol{\theta}|\Sigma, \gamma \sim N(\boldsymbol{\mu}, \bar{V}(\gamma) \otimes \Sigma)$, the entries of $\boldsymbol{\mu}$ are set to replicate a random walk for real GDP and the GDP deflator, and a white noise process for the remaining variables. We then set $\bar{V}(\gamma)$ as discussed in Section 2.2.

We set the prior $\tilde{p}(c)$ equal to a Normal distribution $N(\mu_c, \sigma_c)$ truncated outside the set (τ_l, τ_h) to avoid extreme values of c, and calibrating σ_c such that 95% of the truncated prior mass is in the set (p_l, p_h). To do this we set μ_c equal to the median of the transition variable (0.0462), (τ_l, τ_h) equal to the 1st and 99th percentiles $(-4.1297, 1.6479)$ and (p_l, p_h) equal to the 30th and 70th percentiles $(-0.2922, 0.5462)$. We use a gamma prior distribution for $\tilde{p}(\gamma)$, setting its hyperparameters to ensure a prior mean and standard deviation of 2 and 3, respectively, and then truncate it to have positive mass only above 0.01 (Gefang and Strachan, 2010). In the sensitivity analysis allowing for heteroskedasticity, we set the prior on ψ to a standard Normal distribution. A candidate draw for $\left(\boldsymbol{\theta}, \Sigma, \gamma, c\right)$ is stored only if the maximum eigenvalue of the companion forms for both Π_a and Π_b associated with $\boldsymbol{\theta}$ are below 1.01, in absolute value.